Optimized gene expression from bacterial chromosome by high-throughput integration and screening

Tatyana E. Saleski[1], Meng Ting Chung[2], David N. Carruthers[1], Azzaya Khasbaatar[1], Katsuo Kurabayashi[2,3], Xiaoxia Nina Lin[1,4]*

[1]Department of Chemical Engineering, University of Michigan, Ann Arbor, MI, USA
[2]Department of Mechanical Engineering, University of Michigan, Ann Arbor, MI, USA
[3]Department of Electrical Engineering and Computer Science, University of Michigan, Ann Arbor, MI, USA
[4]Department of Biomedical Engineering, University of Michigan, Ann Arbor, MI, USA

*Corresponding author. Email: ninalin@umich.edu

Keywords: genome-scale engineering, chromosomal integration, gene expression, high-throughput screening, *Escherichia coli*

Running title: Optimizing expression by Tn5 integration

Abstract

Chromosomal integration of recombinant genes is desirable compared to expression from plasmids due to increased stability, reduced cell-to-cell variability, and the elimination of antibiotics for plasmid maintenance. Here, we present a new approach for tuning pathway gene expression levels via random integrations followed by high-throughput screening. We demonstrate multiplexed pathway gene integration and optimization of expression levels for isobutanol production in Escherichia coli. The integrated strains could, with significantly lower expression levels than observed from multicopy plasmid-based expression, produce high titers (up to 10.0 ± 0.9 g/L isobutanol in 48 h) and yields (up to 69 % of the theoretical maximum). Close examination of pathway gene expression in the top-performing, as well as other isolates, reveals the complexity of cellular metabolism and regulation, underscoring the need for precise optimization while integrating pathway genes into the chromosome. This new method for multiplexed pathway gene integration and expression optimization could be readily extended and applied to a wide range of pathways and chassis to create robust and efficient production strains.

Introduction

Microbial biosynthesis is a sustainable, high-specificity approach to achieving chemical conversions with the potential to produce a vast assortment of pharmaceutical, fuel, and commodity chemicals. Development of a high-performing production strain for a desired molecule generally requires tuning the expression levels of native and/or heterologous genes in hosts such as *Escherichia coli*[1, 2]. While extrachromosomal multicopy plasmids offer a convenient method for rapidly prototyping different expression levels of genes, and the majority of metabolic engineering efforts utilize them, they have several challenging attributes. First and foremost, plasmids are unstable, which has been extensively documented and studied due to the importance of plasmids in biotechnology[3-9]. Cells with plasmids can suffer from both structural instability, in which the plasmid is still carried but mutations inactivate the gene of interest, as well as segregational instability, in which some cells no longer carry the plasmid[10, 11]. Plasmids can also multimerize, which decreases their stability[12, 13]. Second, plasmid copy number can vary between cells within stable populations, and the degree of variation is often not well characterized even for widely used plasmids[14]. Münch *et al.* presented a striking example of cellular heterogeneity during plasmid-based production of recombinant protein in *Bacillus megaterium*. The authors found that, under the condition of strong selective pressure, 30% of the cell population was in a low-producing state[15]. This heterogeneity was found to be a result of asymmetric plasmid distribution. Moreover, plasmids require a selective pressure, which is typically an antibiotic molecule. The addition of antibiotics raises the process costs and puts additional stress on the cells. It is also not always effective. Kanamycin, for example, can become an ineffective selection agent at high phosphate concentrations[16]. Furthermore, due to the rise of multidrug-resistant pathogens, there is strong motivation to reduce antibiotic usage[17].

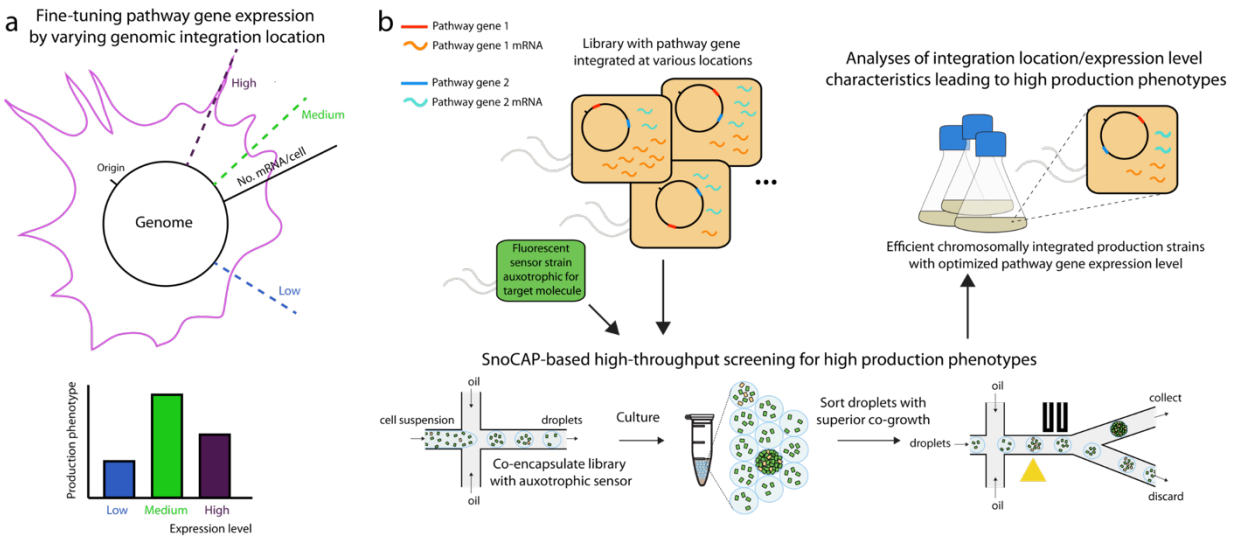

Figure 1. New approach for optimizing chromosomal heterologous gene expression by position-dependent expression variation. (a) Gene expression level varies depending on distance from the origin of replication, as well as other position-dependent factors. The expression level of a heterologous pathway gene affects the production phenotype of a strain. Increasing the expression level of a gene can improve production level by increasing flux through the pathway, but avoiding too high of an expression level can also provide benefits such as reducing the buildup of toxic intermediates and avoiding unnecessary metabolic burden. (b) Libraries with pathway genes integrated into various genome locations can be screened for production phenotype using syntrophic co-culture amplification of production (SnoCAP)[18]. A cross-feeding co-culture is configured consisting of a secretor strain producing a target molecule and a sensor strain that cannot synthesize the target molecule autonomously. The co-cultures are compartmentalized in microfluidic droplets, incubated to allow co-growth, and then sorted to identify the strains with high production phenotypes.

Chromosomal integration and expression of genes is an effective way for circumventing the issues of plasmids described above. This alternative approach leads to more stable and robust production strains, as well as reduce leakiness of inducible expression systems, which are highly desirable traits for large-scale and long-term production processes[19, 20]. Mairhofer *et al.* compared a plasmid-based to a chromosomally integrated expression system and found a much more significant stress response in the plasmid-carrying strain. They identified massive overtranscription of the plasmid-based gene of interest, leading to diversion of ribosomes and other cellular resources, as the source of the metabolic burden[21].

Significant research efforts have been devoted to developing techniques for synthetic integrations[22-24]. The two tools most widely used in prokaryotes for targeted integration of constructs are homologous recombination by expression of the RecET or λ-Red proteins[25-27] and site-specific recombination[28]. Homologous recombination enables rapid integration into a desired locus but is only suitable for constructs up to several kilobases in size and that do not carry significant homology to another part of the genome. Site-specific recombination is less restrictive of the input construct but requires a bacterial chromosomal attachment site (*attB*) in the recipient genome, which can place limitations on the location of integration. This site restrictiveness can be circumvented by first inserting the attachment site into the target locus[29]. Yet, despite the recognized desirability of chromosomal integration and the genetic tools available for targeted integrations, achieving suitable gene expression levels from the chromosome for optimal production remains far more challenging than manipulation of gene expression via plasmids, even in well-studied organisms such as *E. coli*[23].

Chromosomal gene expression levels are dependent on multiple determinants. Similar to plasmid-based gene expression, elements of the genetic construct, including promoters, ribosome-binding sites (RBSs), and enhancers or activators, can be exploited to optimize heterologous gene expression. Yet, single-copy expression from the chromosome is generally weaker than that of the same construct from a multi-copy plasmid. Subsequently, there have been efforts in increasing the copy number on the chromosome to achieve high-level gene expression[19, 30].

Another determinant affecting the expression level from the chromosome strongly is the location of gene integration. It has long been known that during exponential growth, genes closer to the origin of replication experience higher gene dosage and, therefore, higher expression levels[31, 32]. More recently it has been shown that other factors of genomic position, such as level

of DNA compaction or proximity to active genes, can also affect gene expression levels[33, 34]. In examining transcription levels of a reporter gene at various sites in the *E. coli* genome, Bryant *et al.*[33] and Scholz *et al.*[34] observed significant differences, up to ~300-fold, in expression across the *E. coli* genome, excluding gene dosage effects. Intriguingly, on the other hand, it has been suggested that in nature, microorganisms may use transposon-mediated transfer of catabolic genes to different genomic locations as a means of fine-tuning their expression levels in order to adapt to new environmental conditions[35]. These findings suggest that modulation of gene integration position can serve as a useful tool for regulating expression levels. For instance, Loeschcke *et al.* have developed an approach for transfer and expression (TREX) of biosynthetic pathways in bacteria utilizing Tn5 transposase to integrate gene clusters into random sites of the chromosomes and, based on resulting colony color intensities, observed a range of production levels for pigmented secondary metabolites[36].

Synthetic biology applications targeting the production of a small molecule often require delicately balanced cellular resource allocation between production and growth, as well as across different genes of the biosynthetic pathway. It remains elusive to predict the optimal gene expression levels for most pathways due to the highly entangled nature of cellular metabolic and regulatory networks. Additionally, for chromosomal gene expression, it is not fully understood yet how the expression of different regions of the chromosome is affected by changes in environmental conditions and whether/how an inserted gene construct will affect the native expression level of the insertion region. It is, therefore, very difficult to rationally decide where to integrate a gene construct. Here, we present a new metabolic engineering strategy for optimizing a pathway's performance by tuning the gene expression through position-dependent expression variation (Fig. 1). In this method, we utilize the Tn5 transposase to randomly integrate pathway genes in the *E.*

coli genome in a multiplexed fashion. We subsequently screen the libraries using syntrophic co-culture amplification of production (SnoCAP), which converts the production phenotype into a readily screenable growth phenotype, as described in our previous work[18]. We demonstrate the approach using a pathway for the production of isobutanol, a promising drop-in biofuel[37].

Efficient isobutanol production in *E. coli* has been achieved by diverting the branched-chain amino acid pathway intermediate 2-ketoisovalerate (2-KIV) for conversion into isobutanol[38] (Fig. 2a). It has been demonstrated that by overexpressing AlsS from *Bacillus subtilis*, IlvC and IlvD from *E. coli* (for the conversion of pyruvate to 2-KIV), and KivD and AdhA from *Lactococcus lactis* (for the conversion of 2-KIV to isobutanol), production can reach up to 84% of the theoretical yield under aerobic batch conditions[38] and titers can reach up to 50.8 g/L under fed-batch conditions[39]. Resolution of a cofactor imbalance has enabled strains to reach 100% of theoretical yield under anaerobic conditions[40], and a recently developed strategy for isobutanol-linked growth selection under anaerobic conditions enabled optimization of pathway expression levels for achieving optimal production rates[41].

While much work with this pathway has relied on plasmid-based expression of pathway genes, there is also interest in developing chromosomally integrated strains. Akita et al. have demonstrated the integration of the five pathway genes into the *E. coli* genome, resulting in a titer of 6.8 g/L and a yield of 55%, without the addition of costly antibiotics or inducers[42]. Bassalo et al. developed a new CRISPR-based method for high-efficiency, single-step chromosomal

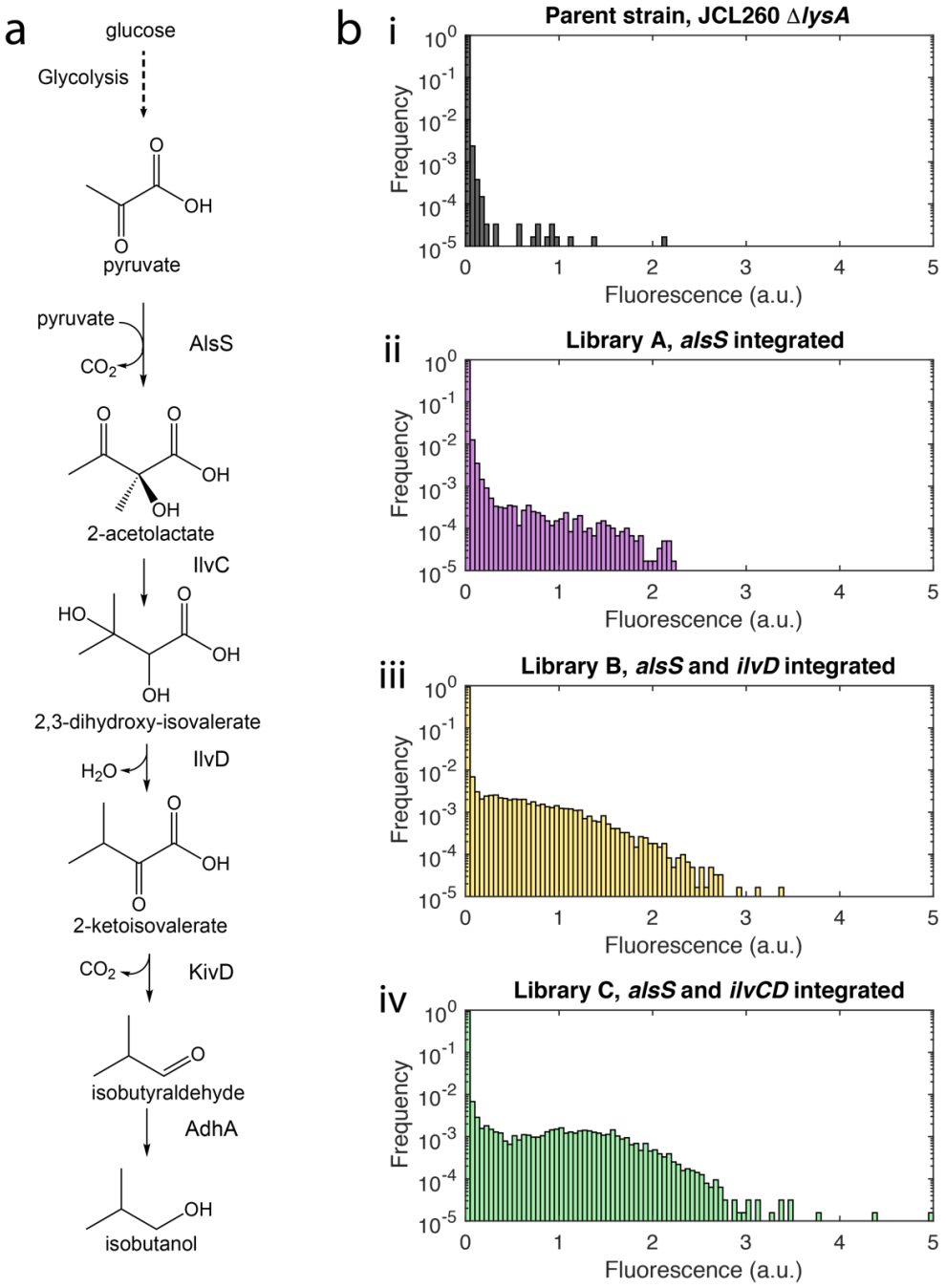

Figure 2. Screening of libraries with *alsS*, *ilvCD* integrated by Tn5. (a) Isobutanol production pathway[43]. Acetolactate synthase (AlsS), ketol-acid reductoisomerase (IlvC), dihydroxy-acid dehydratase (IlvD), 2-ketoacid decarboxylase (KivD), and alcohol dehydrogenase (AdhA). (b) Histograms comparing fluorescence signal from droplets containing co-cultures of fluorescent sensor strain with unlabeled secretor strain libraries, compared to the parent strain (i). The libraries were constructed by Tn5 integration of *kan*-PLlacO1-*alsS* (Library A) (ii), PLlacO1-*ilvD* followed by *kan*-PLlacO1-*alsS* (Library B) (iii), or PLlacO1-*ilvCD* followed by *kan*-PLlacO1-*alsS* (Library C) (iv). Each histogram comprises data from the analysis of about 60,000 droplets.

integration in *E. coli* and, in demonstrating its significant size capability, reported integration of the entire five-gene isobutanol pathway into a single chromosomal locus, leading to 2.2 g/L isobutanol from 85 g/L glucose[44].

In the present work, we demonstrate that the generation of libraries with isobutanol pathway genes integrated into diverse chromosomal positions combined with high-throughput production phenotype screening is an effective means of generating high-performance chromosomally integrated strains. *E. coli* strains resulted from this study show, to our knowledge, the highest titers and yields yet reported with the isobutanol pathway genes expressed exclusively from the chromosome.

Results

Construction and screening of Tn5 integration libraries

We began by constructing an integration library (Library A) in which a construct containing the *alsS* gene, under the PLlacO1 promoter, and a kanamycin-resistance gene, was integrated by Tn5 transposase into the genome of JCL260 Δ*lysA* in random locations. This integration was accomplished by PCR amplification of the *alsS* and *kan* genes with primers that add the 19 bp mosaic end sequence recognized by Tn5, followed by reaction with Tn5, and electroporation of cells with the resulting transposome. The approximate library size was determined to be ~$3.4 \pm 1.1 \times 10^4$ by plating a small quantity of the cells after a short initial recovery on kanamycin-containing plates. The remainder of the library was subjected to a liquid kanamycin selection and subsequently screened by SnoCAP. In the SnoCAP approach, the library is co-encapsulated in water-in-oil microdroplets with a fluorescent sensor strain that is auxotrophic for the target molecule of interest. The library itself is also auxotrophic for a second, orthogonal molecule supplied by the sensor strain. This cross-feeding configuration enables conversion of the

production phenotype into a fluorescent output, amplified by co-culture growth[18]. To screen for overproducers of 2-ketoisovalerate, an intermediate in the isobutanol pathway (Fig. 2a), we use a sensor strain with a deletion of the *ilvD* gene (encoding a dihydroxy-acid dehydratase), which prevents it from converting 2,3-dihydroxy-isovalerate into 2-ketoisovalerate. This strain can only grow when it is provided an exogenous source of 2-ketoisovalerate and/or branched-chain amino acids. We demonstrated in our previous work that this sensor can be used to identify strains that overproduce isobutanol once the genes responsible for converting 2-ketoisovalerate to isobutanol (*kivD* and *adhA*) are added[18].

The cells are loaded at densities such that all droplets contain sensor cells, and most droplets contain either zero or one secretor cell from the integration library. Thus, the co-growth phenotype of individual library members may be assessed. The microdroplet compartmentalization format enables high-throughput screening and isolation of droplets of interest by fluorescence-activated droplet sorting (FADS)[45].

Fluorescence measurement of the droplets containing Library A members co-cultured with fluorescent sensor strain showed that more droplets exhibited growth than when droplets contained co-cultures with the parent strain, which lacks the *alsS* gene, as the secretor (Fig. 2bi,ii). We sorted the top 0.3% of droplets with the highest fluorescence signal and plated the pool of droplets on plates with tetracycline to select for the secretor cells and against the sensor cells (the parent strain of the secretor strain libraries contains a tetracycline resistance gene in the F plasmid, whereas the sensor strain is tetracycline sensitive). We assessed the isobutanol production of colonies obtained from the screening and colonies from the unscreened library, after transformation with pSA65. Of 12 colonies chosen at random from the unscreened library, 10 produced less than 0.1 g/L after 48 h of fermentation (similar to the parent strain), one produced 0.4 g/L and the other produced 2.1

g/L (Fig. S1). Of 10 random colonies tested from the sorted pool, all produced more than 2.5 g/L (Fig. S1), which demonstrates the effectiveness of the screening method.

Because transposome integration efficiency is expected to decrease with increasing construct length, we chose to use the *ilvCD* genes as a selection marker by utilizing their essentiality in minimal medium, rather than an antibiotic. We deleted *ilvD* from JCL260 Δ*lysA* and integrated constructs containing either both genes *ilvC* and *ilvD* behind the PLlacO1 promoter or only the *ilvD* gene behind the same promoter. We performed the selection in minimal M9 medium supplemented with IPTG and lysine, with a small amount plated on solid medium for library size estimation and the rest of the library undergoing liquid selection. These integrations produced far smaller library sizes than the *alsS* integration construct (~10^2 integrants per electroporation vs. ~10^4 for *alsS*). We determined that this was due to substantially lower survival rates in the minimal medium selection compared to kanamycin selection in LB medium. For example, when JCL260 Δ*lysA* Δ*ilvD* was transformed with pSA69, the number of colonies recovered on LB plates with kanamycin was two orders of magnitude higher than the number on M9 plates with lysine and IPTG. Further optimization of the recovery method, such as supplementation of additional nutrients or seeding of small quantities of the branched-chain amino acids may help improve this lower-than-expected recovery rate and increase the diversity of the library. It should also be noted that the survival rate on the solid medium may be different from that in the liquid medium, which could contribute to inaccuracies in estimated library size. Deep sequencing of the library could be used to quantify the library size more accurately. We repeated the *ilvD* and *ilvCD* library generation several times and pooled the integrants, resulting in libraries of each of size ~500-1,000 (estimated from agar plating). We then integrated the *alsS-kan* transposome into these libraries, which led to ~$1.1 \pm 0.5 \times 10^4$ integrants for the *ilvD/alsS* library (Library B) and $1.3 \pm 0.3 \times 10^4$

for the *ilvCD/alsS* library (Library C). Therefore, the library should consist of about 10 different *alsS* integration locations for each *ilvD* or *ilvCD* integration library member.

When assessed by the SnoCAP method (library members co-encapsulated in microdroplets with sensor strain, incubated for co-growth, and the fluorescence of the droplets measured), the *ilvD* and *ilvCD* integration libraries displayed no measurable co-growth. The *ilvD/alsS* and *ilvCD/alsS* libraries (Libraries B and C, respectively) on the other hand, showed substantially more fluorescence compared to the library with only *alsS* integrated (Library A) (Fig. 2b).

We sorted the droplets containing co-cultures of these libraries and recovered colonies from the pools of droplets, and then transformed the isolates with plasmid pSA65 (encoding KivD and AdhA) and assessed isobutanol production. The best isolates were found to be from Library C. In particular, an isolate termed C7 reached a production level closely approaching that of the strain expressing the pathway genes from a plasmid (Fig. 3a). With the goal of understanding the relationship between production phenotype and genome-position-dependent gene expression, we decided to examine isolates displaying a spectrum of production levels and hence also isolated several library members with low- and intermediate-production levels (Fig. 3a) by screening the random libraries via the microplate format of SnoCAP.

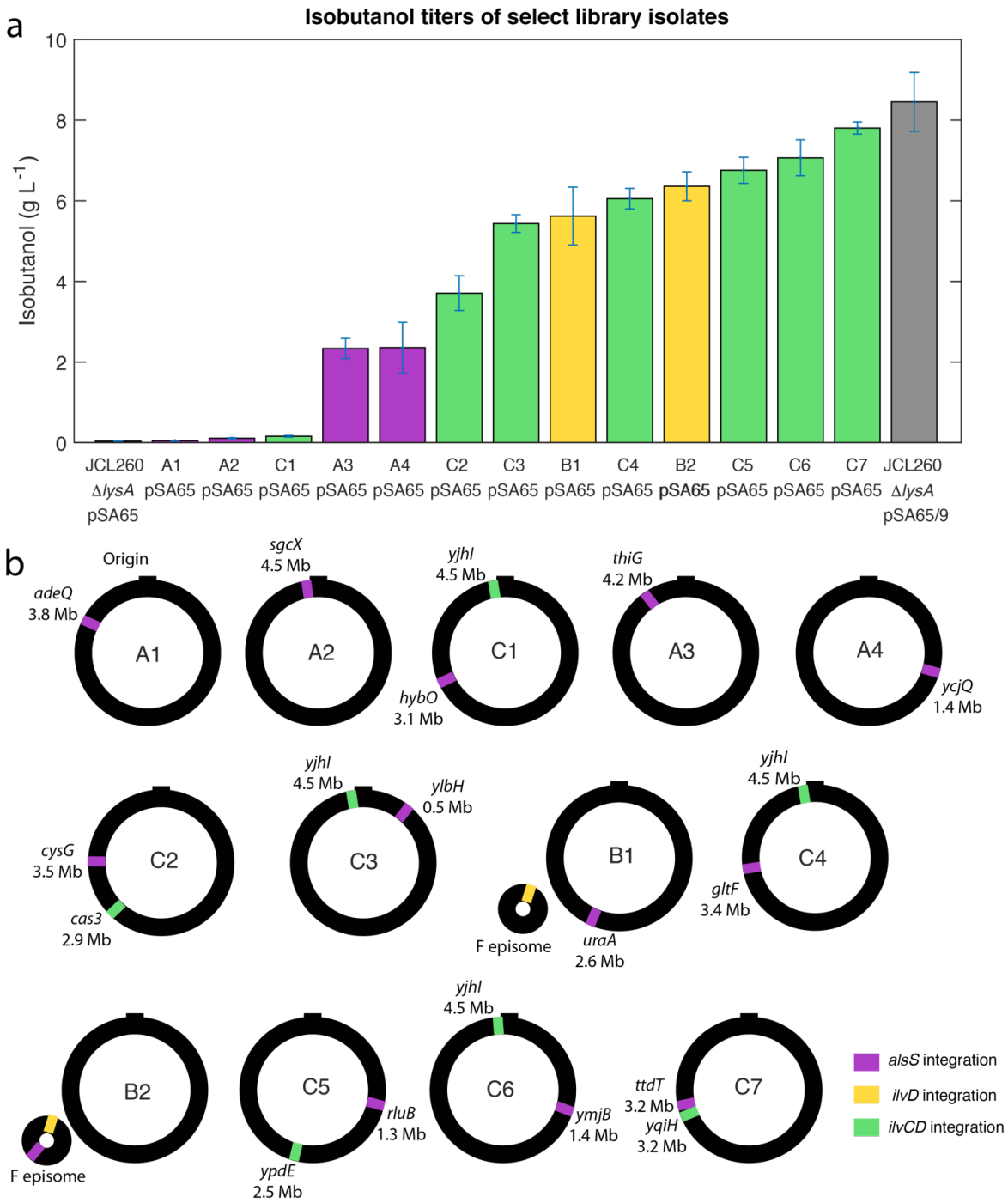

Figure 3. Production phenotype and gene insertion locations of select library isolates. a) Isobutanol titers of library isolates displaying various production levels. Samples were taken 48 h after induction of the pathway genes. Each of the isolates has been transformed with the pSA65 plasmid expressing *kivD* and *adhA*. Parent strain JCL260 Δ*lysA* and JCL260 Δ*lysA* pSA69, which expresses the *alsS* and *ilvCD* genes from the pSA69 plasmid, each also carrying pSA65, are included for comparison. Error bars represent the standard deviation of three biological replicates. (b) Gene integration locations of the isolates whose production is shown in (a).

Table 1. Gene integration locations.

Strain	Integration construct	Gene	Coordinate in gene of integration location	Direction relative to gene integrated into
A1	*kan-alsS*	*adeQ*	436	Opposite
A2	*kan-alsS*	*sgcX*	111	Opposite
C1	*kan-alsS*	*hybO*	812	Opposite
C1	*ilvCD*	*yjhI*	231	Same
A3	*kan-alsS*	*thiG*	73	Same
A4	*kan-alsS*	*ycjQ*	364	Opposite
C2	*kan-alsS*	*cysG*	563	Opposite
C2	*ilvCD*	*cas3*	1,387	Same
C3	*kan-alsS*	*ylbH*	205	Opposite
C3	*ilvCD*	*yjhI*	231	Same
B1	*kan-alsS*	*uraA*	659	Same
B1	*ilvD*	gamma delta (Tn1000) region of F plasmid	426 bp into *tnpX*	Same as *tnpX*
C4	*kan-alsS*	*gltF*	313	Opposite
C4	*ilvCD*	*yjhI*	231	Same
B2	*kan-alsS*	Upstream of *psiB* on F plasmid	218 bp upstream	Opposite of *psiB*
B2	*ilvD*	gamma delta (Tn1000) region of F plasmid	426 bp into *tnpX*	Same as *tnpX*
C5	*kan-alsS*	*rluB*	794	Same
C5	*ilvCD*	*ypdE*	76	Same
C6	*kan-alsS*	*ymjB*	7	Opposite
C6	*ilvCD*	*yjhI*	231	Same
C7	*kan-alsS*	*ttdT*	114	Opposite
C7	*ilvCD*	*yqiH*	512	Opposite

Genomic characterization of select library isolates

We then determined the insertion site locations in the selected isolates by transposon footprinting, which also allowed us to eliminate duplicates of the same genotype and resulted in a panel of unique isolates reported in this work. The integration locations were found to spread across the whole genome (Fig. 3b, Table 1). Intriguingly, each of the Library B isolates, B1 and B2, contains integrations into the F plasmid, a large (~100 kb) plasmid previously introduced to the base strain to supply *lacIq* for increased levels of the lac repressor. The F plasmid is stably maintained in *E. coli* due to an active partitioning system[46, 47], and it has been suggested that placing heterologous genes on the F plasmid may provide improved stability over the use of other plasmids[48, 49]. It is also interesting that in isolate B2, both integration constructs are found to be located on the F plasmid.

Another noteworthy observation was that two of the disrupted genes, *ttdT* and *yjhI*, are genes in which we had found mutations during previous work in which we chemically mutagenized JCL260 Δ*lysA*, screened for improved isobutanol productivity and resequenced the genome of an improved isolate[18]. In that study, the strain with the best production level possessed a mutation in the *ttdT* gene leading to an alanine to valine substitution at position 390 of the protein, and *yjhI* contained a mutation leading to a premature stop codon after the first 187 amino acids (of 227).

In the best isolate, C7, the two integration constructs landed in *ttdT* and *yqiH,* two genes that are quite close to each other, separated by only 18 kb (Fig. 4a). To understand the genetic mechanism underlying the high-production phenotype of this isolate, we investigated whether the gene disruptions, besides expression of the inserted genes, can contribute to increased production. We knocked out the *ttdT* and *yqiH* genes individually in a base strain JCL260 Δ*lysA*.

Into each deletion strain, we then integrated *alsS* in the *yghX* site, which we have previously seen leads to an intermediate production level (and therefore should allow us to see changes in the production that may not be visible in a very low- or high-producing background strain). After transformation of the strains with pSA65, we tested isobutanol production and found that the *ttdT* deletion strain produces more isobutanol to a statistically significant degree compared to the strain possessing wildtype *ttdT*, whereas *yqiH* deletion does not have a statistically significant effect (Fig. 4b).

Characterization of pathway gene transcript levels in select library isolates

We next analyzed gene expression in various isolates of different production levels. We harvested the RNA during exponential growth (3 h after induction of the pathway genes with IPTG), reverse transcribed it and analyzed gene expression of *alsS*, *ilvC*, and *ilvD* by quantitative real-time PCR (qPCR) (Fig. 5b), while meanwhile continuing to grow the cultures and measuring their isobutanol production 48 h following induction (Fig. 5a). It was found that the pathway gene expression in each of the integrated strains was, as expected, far lower than that in the strain expressing *alsS/ilvCD* from the pSA69 plasmid (Fig. 5b, note the log-scale of the Y-axis). Yet, with significantly lower expression levels of all three genes, the top isolate C7 showed a production level very close to that of the plasmid strain.

With the gene transcript profiles of these isolates, we sought to elucidate the relationship between gene expression and production phenotype. Some interesting observations were made. For instance, predictably, isolates from Library A have lower expression of *ilvD* than isolates from Libraries B and C. However, *ilvC* can be seen to be upregulated even in certain strains that do not have *ilvC* insertion (A3, A4, B1). These strains have high expression levels of *alsS*,

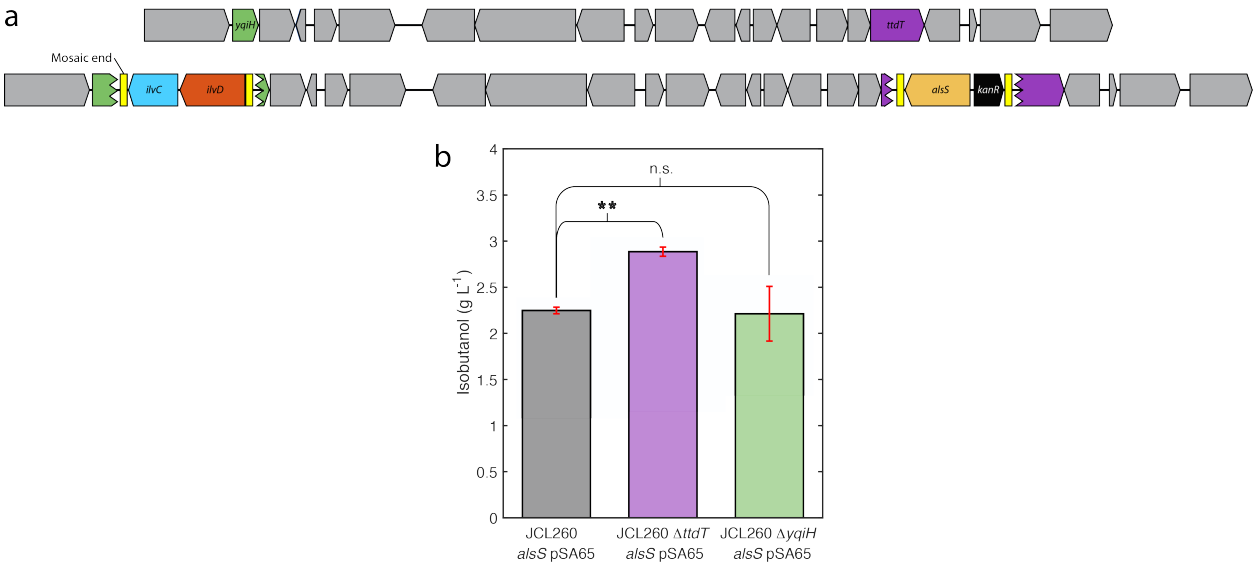

Figure 4. Examination of the effect of gene disruptions incurred during the Tn5 integrations in the C7 isolate on production phenotype. (a) Schematic of the locations of *kan*-PLlacO1-*alsS* and PLlacO1-*ilvCD* integrations in isolate C7, determined by transposon footprinting. (b) Isobutanol production of strains with *alsS* genomically integrated, containing plasmid pSA65 encoding *kivD* and *adhA*, and possessing deletions of either *ttdT* or *yqiH*, compared to the same strain without either deletion. Data shown are from samples taken 72 h after induction with IPTG. Error bars in (b) represent the standard deviation of three biological replicates. **: significant difference at *P*-value < 0.0001 using an unpaired t-test. n.s.: not significant.

so it is likely that accumulation of acetolactate, the product of AlsS, is inducing higher transcription levels of *ilvC* in these strains, as has been previously observed[50]. We also noted that there was a similar trend of relative expression levels within the pathway (*ilvC* highest followed by *alsS* and then *ilvD*) in the medium- and high-production isolates (A3, A4, B1, C5, C6, C7), which was also seen in the plasmid strain. Most strikingly, however, there was a lack of a clear relationship between gene expression levels and production. In particular, despite the fact that the top four analyzed isolates (B1, C5, C6, C7) all showed a much higher expression of *ilvD* than the next two, substantially lower-producing, isolates (A3, A4), it is not clear what causes C7 to perform significantly better in production than the other three isolates or why differences in gene expression levels across these three do not lead to differences in production.

Examination of the low-production isolates also provided some insights. Low expression, especially of *alsS*, appeared to have caused low production in A1 and A2. Intriguingly, another isolate C1 showed much higher expression of the three pathway genes but did not exhibit increased production. We suspect that this could be due to deleterious effects of disrupting the genes in the integration sites or to poor pathway balancing. The above findings underscore the highly entangled and complex nature of optimizing pathway gene expression for a specific production phenotype.

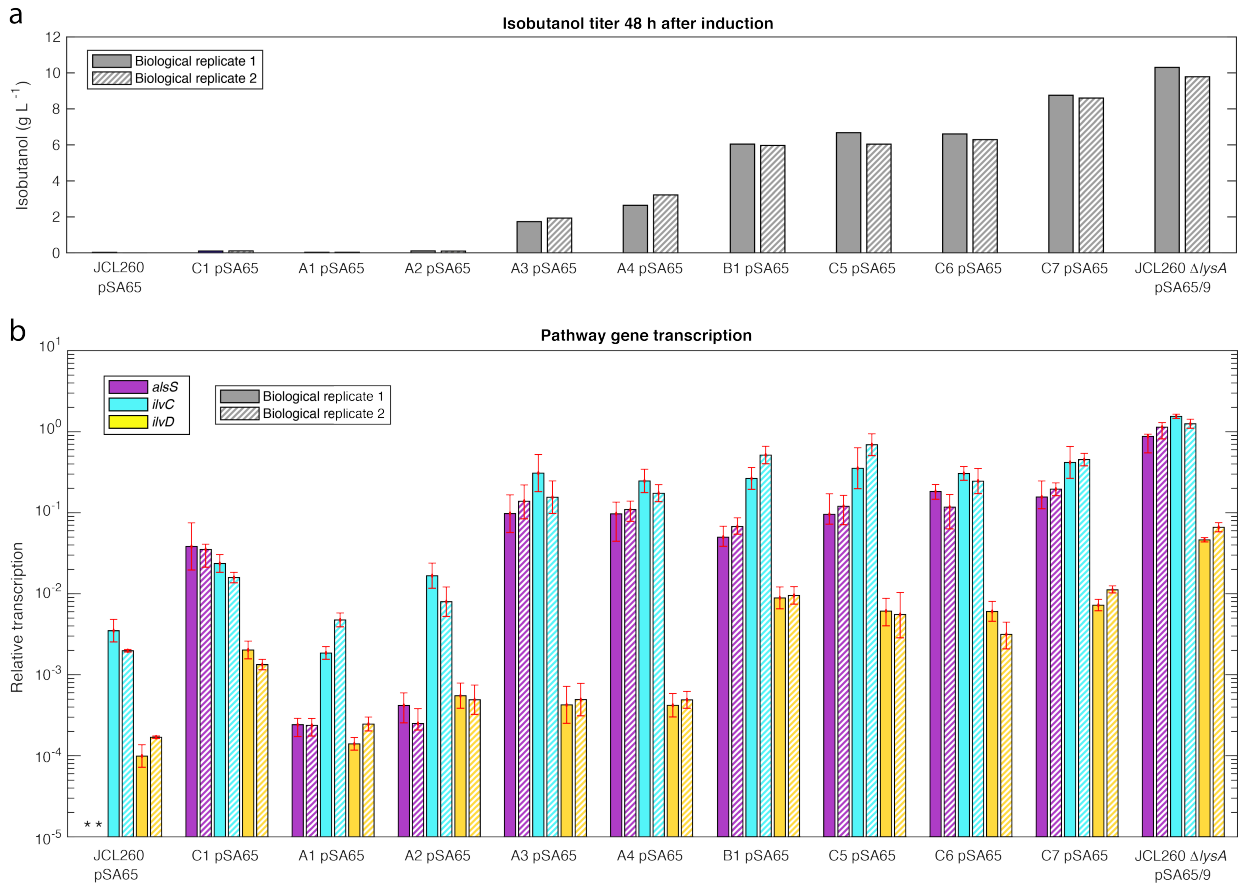

Figure 5. Investigation of the expression levels underlying production phenotypes of select library isolates. (a) Isobutanol production 48 h after induction of the cultures from which the RNA had been harvested. (b) Expression levels of *alsS*, *ilvC*, and *ilvD*, relative to *alsS* level in JCL260 Δ*lysA* pSA65/9. RNA was harvested 3 h after induction with IPTG. * indicates no amplification during the qPCR reaction. Data from two biological replicate cultures of each strain are shown. Error bars in (b) represent the standard deviation of three technical replicates.

Construction and characterization of strains with all isobutanol pathway genes integrated

To generate strains in which the entire isobutanol pathway is expressed from the chromosome, we integrated the *adhA* and *kivD* genes into isolate C7. Because in the work of Atsumi et al.[38], these genes were expressed from a high-copy-number plasmid, we expected that high expression levels would be needed to achieve suitable production levels. We utilized the chemically inducible chromosomal evolution (CIChE) method of Tyo et al.[19] for generating multiple copies of a gene in a single integration site. In this approach, the genes of interest are placed alongside an antibiotic marker between 1 kb homology regions that enables RecA-mediated recombination to increase the copy number in response to increases in antibiotic concentration (Fig. 6a). If desired, the copy number can be stabilized by deletion of the *recA* gene. We integrated the CIChE construct containing *kivD, adhA,* and the chloramphenicol resistance gene *cat* into the *aslB* site, a high-expression site[33]. Before integrating this construct into C7, we also integrated it into a clean background strain, passaged it into higher chloramphenicol concentrations, deleted the *recA* gene to stabilize the copy number, and measured the construct copy number by qPCR of the *cat* gene (Fig. S2). These measurements demonstrated that the copy number can be increased up to about 60 copies by this method. We tested the production of the C7 strain with the CIChE *kivD-adhA* integration (C7 CIChE) in 50 μg/mL chloramphenicol and found that it produced more isobutanol than a strain with just a single copy of *kivD* and *adhA* integrated into the same site. Further increasing the chloramphenicol concentration did not lead to an increase in production level. The fully integrated strain C7 CIChE performed very similarly to C7 pSA65, generating isobutanol titers and yields lower than but close to those of the double plasmid strain (Fig. 6b,c). Noticeably, on

the other hand, this strain reached higher cell densities than the plasmid-bearing strains (Fig. 6d).

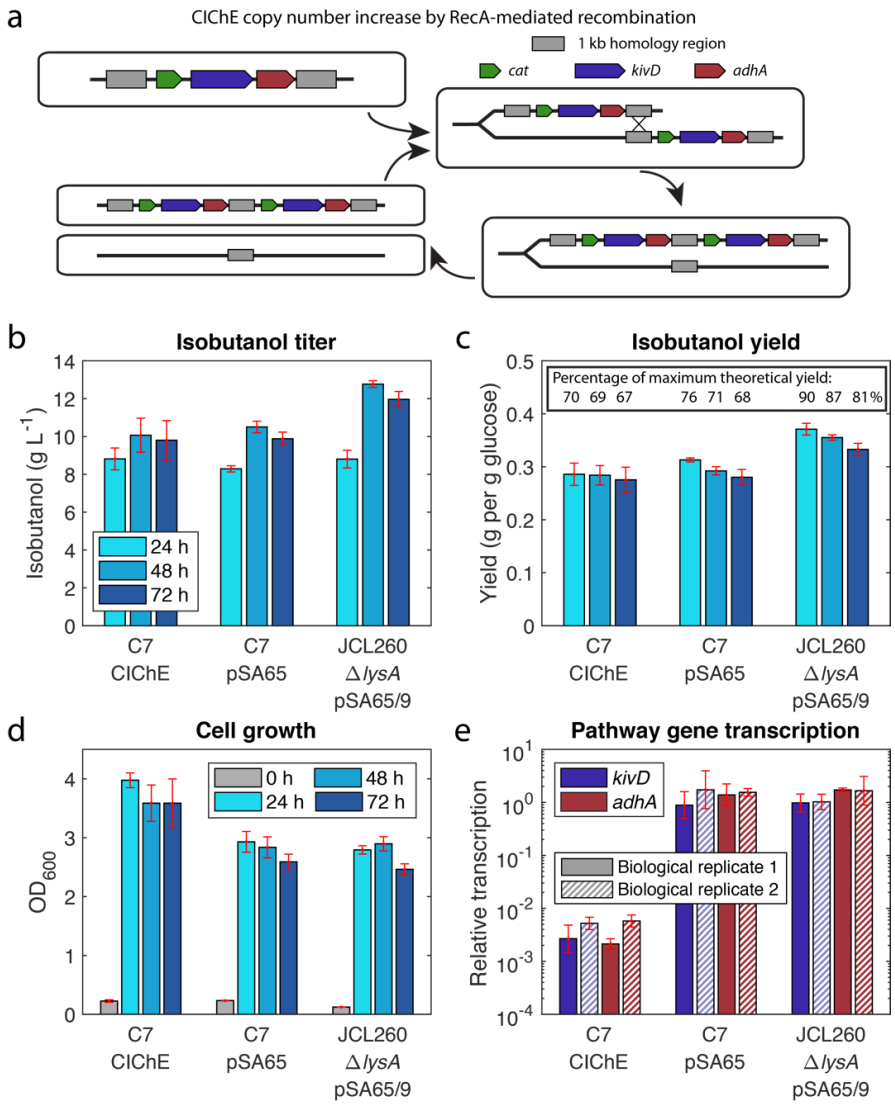

Figure 6. Fully integrated strain performance compared to strains with plasmids. (a) Overview of the chemically inducible chromosomal evolution (CIChE) approach for gene copy number amplification by RecA-mediated homologous recombination. (b) Isobutanol production by strain C7 after integration of *kivD-adhA* by CIChE, compared to C7 pSA65, which has *alsS* and *ilvCD* integrated into the genome and *kivD* and *adhA* expressed from the pSA65 plasmid, and JCL260 Δ*lysA* pSA65/9 which has all five pathway genes expressed from the plasmids. Error bars represent the standard deviation of three biological replicates. (c) The yields of the cultures based on glucose consumed. The inset in (c) lists the percentage of the theoretical maximum yield (0.41 g isobutanol per gram glucose). (d) The cell densities of the production cultures whose performance is displayed in (b) and (c). (e) Expression levels of *kivD* and *adhA*, relative to *kivD* level in JCL260 Δ*lysA* pSA65/9. RNA was harvested 4 h after induction with IPTG. Error bars represent the standard deviation of three technical replicates.

We also examined the *kivD* and *adhA* transcript levels in the fully integrated strain compared to the plasmid-bearing strains. As expected, they are lower in C7 CIChE. Quite surprisingly, however, we found them to be dramatically lower (more than two orders of magnitude) in the integrated strain compared to those in both strains expressing *kivD*/*adhA* from plasmid pSA65 (Fig. 6e). This indicates that the expression levels rendered by the high-copy number plasmid are unnecessarily high. This result further demonstrates that it is feasible to achieve high production through significantly lower pathway gene expression from the chromosome.

Discussion

Synthetic biology is making strides toward solutions to some of society's most pressing problems. Managing the issue of stability and cell-to-cell variability, however, still present major hurdles to overcome[51]. Moving gene expression from plasmids to the genome can improve stability and reduce cell-to-cell variability. Yet it remains challenging to predict the expression levels that will be optimal for production and achievable by genomic integration into a particular site. Moreover, the interplay between gene integration and surrounding regions on the chromosome, such as gene disruption and expression level perturbation, could cause complications and additional difficulty in optimizing pathway expression.

The new approach of screening of random integration libraries developed in the present study circumvents the challenges of rational selection of integration locations. The transposon integration method is versatile and applicable to almost any gene of interest. Since the integration is not homology-based (unlike lambda Red recombinase-based integration, for example) it can even be used on genes that are already natively present in the genome at suboptimal expression

levels, such as *ilvC* in the work presented here. The high-throughput screening enables large libraries to be assessed, making it feasible to multiplex integrations of different pathway genes.

The application of the approach to the isobutanol pathway demonstrates its effectiveness and has generated a strain with, to our knowledge, the highest production titers yet reported in the literature for an *E. coli* strain with the isobutanol pathway chromosomally integrated. This chromosomally integrated strain, which achieves production performance close to that of a plasmid-based strain with substantially higher pathway gene expression levels, also provides a striking example of how plasmid-based expression can be unnecessarily high. We expect that the lower expression levels in the integrants may be beneficial under more stressful conditions, alleviating unnecessary demands on the cell. Even under the relatively optimal fermentation conditions used here, the integrated strains reach higher cell densities than the plasmid-bearing strains, demonstrating the reduced burden on these strains. Furthermore, we found this approach can take advantage of beneficial gene deletions that may not be obvious choices for targeted deletion to yield optimal production.

Our results also open intriguing questions for future investigation. For example, it will be interesting to learn what factors can lead a strain with relatively high expression levels of all pathway genes, such as C1, to have such a substantially lower production level than other strains. Also intriguing is the fact that, in the highest-producing isolate we identified, C7, as well as in another high-producing isolate, B2, the gene integrations are located quite close to each other. Whether this is due to that region having optimal expression levels, or the proximity of the constructs provides some benefit, or it is merely by chance is not clear. Deep-sequencing pools of droplets collected from different degrees of screening stringency could provide further insights

into whether proximity leads to improvements in production. Additionally, studies with targeted integrations could be designed to address this question.

Methods

Strains and plasmids

Strains and plasmids used are listed in Table S1. JCL260, pSA65, and pSA69[38, 43] were provided by James Liao, UCLA. Keio strains were obtained from the *E. coli* genetic stock center (CGSC, http://cgsc2.biology.yale.edu/). Plasmid pSAS31 was constructed by Scott Scholz[34].

Media

M9IPG, consisting of M9 salts (47.8 mM Na_2HPO_4, 22.0 mM KH_2PO_4, 8.55 mM NaCl, 9.35 mM NH_4Cl, 1 mM $MgSO_4$, 0.3 mM $CaCl_2$), micronutrients (2.91 nM $(NH_4)_2MoO_4$, 401.1 nM H_3BO_3, 30.3 nM $CoCl_2$, 9.61 nM $CuSO_4$, 51.4 nM $MnCl_2$, 6.1 nM $ZnSO_4$, 0.01 mM $FeSO_4$), thiamine HCl (3.32 μM) and dextrose (D-glucose) at the stated concentrations, was used for all culturing experiments. For SnoCAP screening, in both microdroplets and microplates, the glucose concentration was 20 g/L, and the medium was supplemented with 3 mM L-isoleucine and 50 μg/mL kanamycin. For isobutanol production monocultures, the medium contained 36 g/L glucose, 5 g/L yeast extract, and no antibiotics were added. Precultures for both screening and production cultures were carried out in LB Lennox with antibiotics appropriate to the strain. When used, antibiotics were supplied at the following concentrations: ampicillin, 100 μg/mL; kanamycin, 50 μg/mL; tetracycline, 10 μg/mL; chloramphenicol, 50 μg/mL.

Transposon integration library construction

All primers and oligonucleotides were ordered from Integrated DNA Technologies, Inc. (IDT) and are listed in Table S2. Each of the integration constructs was amplified by PCR from the pSA69 plasmid, adding the Tn5 mosaic ends. Primers phosph_transp_kan_for and phosph_transp_alsS_rev were used to amplify *kan*-PLlacO1-*alsS* (for Libraries A, B and C). phosph_transp_PLlacO1_ilvC_for and phosph_transp_ilvD_rev were used to amplify *ilvCD* and introduce a PLlacO1 promoter in front of *ilvC* (for Library C). phosph_transp_PLlacO1_ilvD_for and phosph_transp_ilvD_rev were used to amplify *ilvD* and introduce a PLlacO1 promoter in front of *ilvD* (for Library B). The linear PCR product was then digested with both DpnI and SpeI enzymes (NEB) to digest the template plasmid, phosphorylated with T4 polynucleotide kinase (NEB), cleaned with a PCR clean-up kit (Qiagen) and eluted in TE buffer. The DNA was then reacted with EZ-Tn5 transposase (Lucigen) according to the manufacturer's instructions. The resulting transposome was then electroporated into the appropriate strain. The cells were recovered for 1.5 hr with 1 mL SOC medium. Then 50 µL was used for dilution and plating on LB with kanamycin plates to assess library size, and the remaining cells were grown to saturation in 100 mL selective medium (either LB with 50 µg/mL kanamycin for *alsS* integration or M9IPG with 0.1 mM IPTG and 3 mM lysine for *ilvCD* or *ilvD* integration). The cells were then frozen in 1 mL aliquots (resuspended in fresh LB with 25% glycerol) and later thawed, washed, and grown in LB with 50 µg/mL kanamycin to prepare them for screening.

SnoCAP screening

SnoCAP screening for 2-KIV production, in droplet and microplate formats, was employed as previously described[18]. Stationary phase cultures in LB were used as the inocula for both formats.

For the microplate format, K12 Δ*ilvD* was used as the sensor strain. M9IPG with 20 g/L glucose, 3 mM isoleucine, and 50 μg/mL kanamycin was used as the medium for all screening co-cultures except for those with the *ilvD* and *ilvCD* libraries (without *alsS* integration), in which the kanamycin was omitted. For the droplet format, K12 Δ*ilvD* pSAS31, which expresses mNeonGreen, was used as the sensor strain. For the droplet sorting assay, the droplet collection device was soaked in a mixture of HFE-7500 oil and water for several days before use in order to improve droplet stability after collection.

Isobutanol fermentations and analysis

Isobutanol cultures were grown similarly to the method previously described[18]. Overnight cultures in LB with antibiotics were diluted 1:100 v/v into 10 mL of M9IPG with 36 g/L glucose, 5 g/L yeast extract, and no antibiotics, in a 125-mL baffled, unvented polypropylene flasks. Cells were grown for 2.5 to 3 h at 37 °C, 250 rpm, followed by induction of the pathway genes with 0.1 mM IPTG. Flasks were then sealed with parafilm and incubated at 30 ^{o}C, 250 rpm. Samples were taken daily to measure cell density and glucose and isobutanol concentrations. Cell density was assessed by measuring the OD_{600} of 200 μL of culture, diluted into the linear range, in a VersaMax microplate reader (Molecular Devices). Glucose and isobutanol were measured by high-performance liquid chromatography (HPLC), as previously described[37]. In the case of cultures in which RNA was harvested, 0.5 mL of cells were harvested with RNAprotect Bacteria Reagent (Qiagen) 3-4 h after induction. RNA was isolated using an RNA mini kit (Qiagen) according to the manufacturer's instructions, including on-column DNAse I digestion. An additional digestion using Turbo DNA-free Kit (Invitrogen) was performed after purification to eliminate any remaining genomic DNA. The RNA was then reverse transcribed to cDNA using MultiScribe

Reverse Transcriptase (Invitrogen) and approximately 400 ng RNA per 20 μL reaction, with the following temperature profile: 25 ^{o}C, 10 min; 37 ^{o}C, 120 min, 85 ^{o}C, 5 min. A no reverse transcriptase control was carried out for each sample and checked with at least one set of primers during the real-time PCR analysis to ensure no significant residual genomic DNA. qPCR was conducted on the cDNA using SYBR Green qPCR Master Mix (Life Technologies) with 20 μL reactions. Reactions were run in 96-well plates on a 7900HT Fast Real-Time PCR Machine (Applied Biosystems), courtesy of the University of Michigan Advanced Genomics Core. The PCR program was as follows: 50 ^{o}C, 2 min; 95 ^{o}C, 10 min; 40 cycles of 95 ^{o}C for 15 s, 60 ^{o}C for 1 min. Primers used are listed in Table S2. Data were analyzed using the $2^{-\Delta\Delta CT}$ method[52].

kivD and adhA integration

A CIChE construct consisting of *kivD, adhA*, and a chloramphenicol resistance gene, *cat*, flanked by 1 kb homology regions, was previously generated and integrated into the *aslB* locus of the NV3r1 strain by λ-Red recombineering[18]. Following the method of Tyo et al.[19], we passaged this strain into increasing concentrations of chloramphenicol and then deleted the *recA* gene by P1 transduction from donor strain BW 26,547 Δ*recA*::*kan* Lambda *recA*+ (obtained from the Coli Genetic Stock Center (CGSC)). Transductants were selected on LB plates with 50 μg/mL kanamycin and the corresponding concentration of chloramphenicol. Genomic DNA was isolated using DNEasy Blood & Tissue Kit (Qiagen), and the copy number was determined by qPCR using the same primers as Tyo et al.[19], with one set for the *cat* gene and one for the *bioA* gene, which is used as a single copy reference gene. Absolute copy number was determined by comparison to a strain with a single copy of *cat* integrated into the *intC* locus (NV3r1 Δ*intC::yfp-cat*). To transfer this integration into the C7 strain generated in this study, we prepared P1 lysates from NV3r1 (*recA*+) with the *kivD, adhA* CIChE construct after growing on 80 μg/mL

chloramphenicol and transduced it into the C7 strain, selecting on LB plates with 40 µg/mL chloramphenicol. Resulting colonies were isolation streaked twice on LB agar with 40 µg/mL chloramphenicol and 0.8 mM sodium citrate.

Acknowledgments

The authors thank Scott Scholz and Peter Freddolino for valuable discussions regarding position-dependent gene expression variation and transposon footprinting, as well as Mark Burns, Lola Eniola-Adefeso, and Chris Barr for the use of equipment in their laboratories. The isobutanol strains and plasmids were originally provided by James Liao (UCLA, formerly). The CIChE construct was constructed using the pTGD plasmid provided by Keith Tyo (Northwestern U). This work was supported by the USDA AFRI NIFA Fellowships Grant Program (Grant no. 2016-67011-24725).

Author Contributions

T.S. and X.N.L. conceived and designed the study. T.S. and A.K. constructed the integration libraries. M.T.C. and K.K. developed the droplet sorting platform. T.S. and M.T.C. performed the library screening. T.S. and D.N.C. characterized the library isolates and fully integrated strains. T.S. and X.N.L. wrote the paper.

Competing Interests

The authors declare no competing interests.

References

1. Woolston, B.M., Edgar, S. & Stephanopoulos, G. Metabolic Engineering: Past and Future. *Annual Review of Chemical and Biomolecular Engineering* **4**, 259–288 (2013).
2. Nielsen, J. & Keasling, Jay D. Engineering cellular metabolism. *Cell* **164**, 1185-1197 (2016).
3. Popov, M., Petrov, S., Nacheva, G., Ivanov, I. & Reichl, U. Effects of a recombinant gene expression on ColE1-like plasmid segregation in *Escherichia coli*. *BMC Biotechnology* **11**, 18 (2011).
4. Bentley, W.E. & Kompala, D.S. Plasmid instability in batch cultures of recombinant bacteria. A laboratory experiment. *Chemical Engineering Education* **24**, 168-172 (1990).
5. Seo, J.H. & Bailey, J.E. A segregated model for plasmid content and product synthesis in unstable binary fission recombinant organisms. *Biotechnology and Bioengineering* **27**, 156-166 (1985).
6. Siegel, R. & Ryu, D.D. Kinetic study of instability of recombinant plasmid pPLc23trpAl in *E. coli* using two-stage continuous culture system. *Biotechnology and Bioengineering* **27**, 28-33 (1985).
7. Imanaka, T., Tsunekawa, H. & Aiba, S. Phenotypic stability of trp operon recombinant plasmids in *Escherichia coli*. *Microbiology* **118**, 253-261 (1980).
8. Imanaka, T. & Aiba, S. A perspective on the application of genetic engineering: stability of recombinant plasmid. *Annals of the New York Academy of Sciences* **369**, 1-14 (1981).
9. Ollis, D. Industrial fermentations with (unstable) recombinant cultures. *Philosophical Transactions of the Royal Society of London. B, Biological Sciences* **297**, 617-629 (1982).
10. Friehs, K.P. "Plasmid copy number and plasmid stability" in New Trends and Developments in Biochemical Engineering 47-82 (Springer, 2004).
11. Ensley, B.D. Stability of recombinant plasmids in industrial microorganisms. *Critical Reviews in Biotechnology* **4**, 263-277 (1986).
12. Summers, D.K. & Sherratt, D.J. Multimerization of high copy number plasmids causes instability: ColE 1 encodes a determinant essential for plasmid monomerization and stability. *Cell* **36**, 1097-1103 (1984).
13. Field, C. & Summers, D. Multicopy plasmid stability: revisiting the dimer catastrophe. *Journal of Theoretical Biology* **291**, 119-127 (2011).
14. Jahn, M., Vorpahl, C., Hübschmann, T., Harms, H. & Müller, S. Copy number variability of expression plasmids determined by cell sorting and Droplet Digital PCR. *Microbial Cell Factories* **15**, 211 (2016).
15. Münch, K.M. et al. Polar fixation of plasmids during recombinant protein production in *Bacillus megaterium* results in population heterogeneity. *Applied and environmental microbiology* **81**, 5976-5986 (2015).
16. Studier, F.W. Protein production by auto-induction in high-density shaking cultures. *Protein Expression and Purification* **41**, 207-234 (2005).
17. Mignon, C., Sodoyer, R. & Werle, B. Antibiotic-free selection in biotherapeutics: now and forever. *Pathogens* **4**, 157-181 (2015).
18. Saleski, T.E. et al. Syntrophic co-culture amplification of production phenotype for high-throughput screening of microbial strain libraries. *Metabolic Engineering* **54**, 232-243 (2019).

19. Tyo, K.E.J., Ajikumar, P.K. & Stephanopoulos, G. Stabilized gene duplication enables long-term selection-free heterologous pathway expression. *Nature Biotechnology* **27**, 760-765 (2009).
20. Striedner, G. et al. Plasmid-free T7-based *Escherichia coli* expression systems. *Biotechnology and Bioengineering* **105**, 786-794 (2010).
21. Mairhofer, J., Scharl, T., Marisch, K., Cserjan-Puschmann, M. & Striedner, G. Comparative transcription profiling and in-depth characterization of plasmid-based and plasmid-free Escherichia coli expression systems under production conditions. *Applied and Environmental Microbiology* **79**, 3802-3812 (2013).
22. Madyagol, M. et al. Gene replacement techniques for Escherichia coli genome modification. *Folia Microbiologica* **56**, 253-263 (2011).
23. Ou, B., Garcia, C., Wang, Y., Zhang, W. & Zhu, G. Techniques for chromosomal integration and expression optimization in Escherichia coli. *Biotechnology and Bioengineering* **115** (2018).
24. Li, L., Liu, X., Wei, K., Lu, Y. & Jiang, W. Synthetic biology approaches for chromosomal integration of genes and pathways in industrial microbial systems. *Biotechnology Advances* **37**, 730-745 (2019).
25. Zhang, Y., Buchholz, F., Muyrers, J.P. & Stewart, A.F. A new logic for DNA engineering using recombination in *Escherichia coli*. *Nature Genetics* **20**, 123-128 (1998).
26. Yu, D., Ellis, H.M., Lee, E.-C., Jenkins, N.A. & Copeland, N.G. An efficient recombination system for chromosome engineering in *Escherichia coli*. *Proceedings of the National Academy of Sciences* **97**, 5978-5983 (2000).
27. Datsenko, K.A. & Wanner, B.L. One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. *Proceedings of the National Academy of Sciences* **97**, 6640-6645 (2000).
28. Grindley, N.D., Whiteson, K.L. & Rice, P.A. Mechanisms of site-specific recombination. *Annual Review of Biochemistry* **75**, 567-605 (2006).
29. Minaeva, N.I. et al. Dual-In/Out strategy for genes integration into bacterial chromosome: a novel approach to step-by-step construction of plasmid-less marker-less recombinant *E. coli* strains with predesigned genome structure. *BMC Biotechnology* **8**, 63 (2008).
30. Yin, J. et al. Effects of chromosomal gene copy number and locations on polyhydroxyalkanoate synthesis by *Escherichia coli* and *Halomonas* sp. *Applied Microbiology and Biotechnology* **99**, 5523-5534 (2015).
31. Beckwith, J.R., Signer, E.R. & Epstein, W. in Cold Spring Harbor symposia on quantitative biology, Vol. 31 393-401 (Cold Spring Harbor Laboratory Press, 1966).
32. Schmid, M.B. & Roth, J.R. Gene location affects expression level in *Salmonella typhimurium*. *Journal of Bacteriology* **169**, 2872-2875 (1987).
33. Bryant, J.A., Sellars, L.E., Busby, S.J. & Lee, D.J. Chromosome position effects on gene expression in *Escherichia coli* K-12. *Nucleic Acids Research* **42**, 11383-11392 (2014).
34. Scholz, S.A. et al. High-resolution mapping of the *Escherichia coli* chromosome reveals positions of high and low transcription. *Cell Systems* **8**, 212-225 (2019).
35. Sousa, C., de Lorenzo, V. & Cebolla, A. Modulation of gene expression through chromosomal positioning in *Escherichia coli*. *Microbiology* **143**, 2071-2078 (1997).

36. Loeschcke, A. et al. TREX: a universal tool for the transfer and expression of biosynthetic pathways in bacteria. *ACS Synthetic Biology* **2**, 22-33 (2012).
37. Chen, C.-T. & Liao, J.C. Frontiers in microbial 1-butanol and isobutanol production. *FEMS Microbiology Letters* **363**, fnw020 (2016).
38. Atsumi, S., Hanai, T. & Liao, J.C. Non-fermentative pathways for synthesis of branched-chain higher alcohols as biofuels. *Nature* **451**, 86 (2008).
39. Baez, A., Cho, K.-M. & Liao, J.C. High-flux isobutanol production using engineered *Escherichia coli*: a bioreactor study with in situ product removal. *Applied Microbiology and Biotechnology* **90**, 1681-1690 (2011).
40. Bastian, S., Liu, X., Meyerowitz, J.T., Snow, C.D., Chen, M.M. and Arnold, F.H. Engineered ketol-acid reductoisomerase and alcohol dehydrogenase enable anaerobic 2-methylpropan-1-ol production at theoretical yield in *Escherichia coli*. *Metabolic Engineering* **13**, 345-352 (2011).
41. Ghosh, I.N., Martien, J., Hebert, A.S., Zhang, Y., Coon, J.J., Amador-Noguez, D. and Landick, R. OptSSeq explores enzyme expression and function landscapes to maximize isobutanol production rate. *Metabolic Engineering* **52**, 324-340 (2019).
42. Akita, H., Nakashima, N. & Hoshino, T. Bacterial production of isobutanol without expensive reagents. *Applied Microbiology and Biotechnology* **99**, 991-999 (2015).
43. Atsumi, S. et al. Engineering the isobutanol biosynthetic pathway in *Escherichia coli* by comparison of three aldehyde reductase/alcohol dehydrogenase genes. *Applied Microbiology and Biotechnology* **85**, 651-657 (2010).
44. Bassalo, M.C. et al. Rapid and efficient one-step metabolic pathway integration in *E. coli*. *ACS Synthetic Biology* **5**, 561-568 (2016).
45. Baret, J.-C. et al. Fluorescence-activated droplet sorting (FADS): efficient microfluidic cell sorting based on enzymatic activity. *Lab on a Chip* **9**, 1850-1858 (2009).
46. Clowes, R.C. Molecular structure of bacterial plasmids. *Bacteriological Reviews* **36**, 361 (1972).
47. Ebersbach, G. & Gerdes, K. Plasmid segregation mechanisms. *Annual Review of Genetics* **39**, 453-479 (2005).
48. Kachroo, A.H., Jayaram, M. & Rowley, P.A. Metabolic engineering without plasmids. *Nature Biotechnology* **27**, 729 (2009).
49. Keasling, J.D. Gene-expression tools for the metabolic engineering of bacteria. *Trends in Biotechnology* **17**, 452-460 (1999).
50. Arfin, S.M., Ratzkin, B. & Umbarger, H. The metabolism of valine and isoleucine in *Escherichia coli.* XVII. The role of induction in the depression of acetohydroxy acid isomeroreductase. *Biochemical and Biophysical Research Communications* **37**, 902-908 (1969).
51. Wang, T. & Dunlop, M.J. Controlling and exploiting cell-to-cell variation in metabolic engineering. *Current Opinion in Biotechnology* **57**, 10-16 (2019).
52. Livak, K.J. & Schmittgen, T.D. Analysis of relative gene expression data using real-time quantitative PCR and the 2(T)(-Delta Delta C) method. *Methods* **25**, 402-408 (2001).